I&EC research
Industrial & Engineering Chemistry Research

Article

pubs.acs.org/IECR

A Two-Scale Pursuit Method for the Tailored Identification and Quantification of Unknown Polymer Additives and Contaminants by ^{1}H NMR

Phuong-Mai Nguyen,[†,‡,§] Cédric Lyathaud,[†] and Olivier Vitrac*[,‡,§]

[†]Chemistry and Physical Chemistry of Materials Division, Laboratoire National de métrologie et d'Essais (LNE), 78197 Trappes Cedex, France

[‡]INRA, UMR 1145 Ingénierie Procédés Aliments, Group "Interactions between Materials and Media in Contact", F-91300, Massy, France

[§]AgroParisTech, UMR 1145 Ingénierie Procédés Aliments, Group "Interactions between Materials and Media in Contact", F-91300, Massy, France

Supporting Information

ABSTRACT: Blind deformulation is an important stake for several industries. This work was motivated by the identification and quantification of contaminants originated from food packaging systems. Many substances originating from plastic materials are indeed suspected to be endocrine disruptors but remain chiefly difficult to separate with spectroscopic techniques. We propose a tailored two-scale pursuit methodology to identify and quantify an arbitrary number of substances from the ^{1}H NMR spectrum of the mixture. Identified substances are included within a library of spectra and can be combined with undocumented ones. To preserve the initial resolution of NMR spectra, peak lines are spanned onto Gaussian kernels so that they can be identified, even when the positions and shapes of multiplets in the mixture are modified within tolerance ranges or when multiplets are overlapping. The deconvolution procedure starts with a crude pairwise search to build a list of likely substances, which is subsequently expanded as nested scenarios. Scenarios are built according to the risk of confusing similar substances. Quantification is carried out on a preference list of substances selected as in a voting system. Using a primary library of 52 substances (corresponding to 279 multiplets and 5620 lines), the reliability and robustness of the method were tested extensively in numerical experiments and by performing the brute-force deformulation of five processed common thermoplastics.

1. INTRODUCTION

Several European regulations, such as the "REACH" regulation (1907/2006/EC)[1] and the regulation of food contact materials (see 1935/2004/EC),[2] enforce strict rules for substances with potential risks of release in the environment or food. Substances are globally managed based on three concepts: positive list (list of authorized ones, excluding others), negative list (substances prohibited or to be phased out), and restrictions of use (authorized with specific conditions). The safety materials intended to be in contact with food are, in particular, covered by the three concepts, sometimes combined together as in composite packaging systems (laminated, printed, etc.). A positive list of additives and monomers (ca. 800 substances) exists for plastic materials, along with maximum acceptable concentration values in food (10/2011/EC[3]). Negative list of enforced are set based on good manufacturing practices encouraged by regulation 2023/2006/EC[4] (e.g., guideline on printing inks[5]) or specific regulations, such as the ban of the foaming agent azodicarbonamide.[6] New health threats caused by substances categorized with endocrine-disrupting properties,[7] with estrogenic activity[8] or with low-dose endocrine effects[9] substantially complicate the situation. Bisphenol A is the most popular example. It is forbidden in infant formula packaging both in the European Union (EU)[10] and in the United States,[11] but for different reasons. In the United States, the ban is based on abandonment and not on safety.[12,13] Conversely, France banned all bisphenols for all food contact applications, starting January 1, 2015.[14] Similar discrepancy appears between "plastic" and REACH regulations: 175 substances are authorized in the first one, whereas 54 are subjected to be phase-out in the second.[15]

Enforcing existing current regulations or best practices is a complicate challenge shared by both industry and authorities. Identifying the absence or presence of substances, as well as their amounts, requires a prior knowledge of sought substances with standard chromatographic techniques.[16] For the sole food contact materials, the considerable work is summarized by two figures: ca. 7000 substances have been listed by the European collaborative project FACET,[17] whereas specific chromatographic methods have been collected by European Laboratory of Reference Materials for only 600 substances.[18] The main goal of this study is to propose a new blind deconvolution procedure of spectra in mixtures and to apply it to ^{1}H NMR identification and quantification of substances from polymer extracts. Several appealing characteristics are proposed for routine identification and concentration estimations in complex

Received: September 12, 2014
Revised: February 24, 2015
Accepted: February 24, 2015
Published: February 24, 2015

ACS Publications © 2015 American Chemical Society

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

mixtures, without sample fractionation. In particular, mixtures can include an unknown number of substances and reference spectra can combine measured and theoretical ones. Without a loss of generality, the developments have been motivated by for the constitution of a large database of the composition of packaging materials sampled on the EU market and to be used along with the concept of rationale design of food packaging systems.[19] The available data are the ^{1}H NMR spectra (experimental and some theoretical ones) of the reference substances organized as a single library. For concision, the presented expandable library comprises only 62 experimental spectra of pure substances, whereas the production one includes more than 300. Each spectrum is encoded without resolution loss onto a set of unitary multiplet functions, which are made shift-invariant. This particular feature enables the use of existing databases, such as the ca. 100 000 Wiley ^{1}H database,[20] and calculated spectra.[21−25] It is worth noticing, that, converse to ^{13}C shifts, proton shifts predictions generate large errors, because of the lack of averaging over many conformers and interactions with solvent.[26]

The main issue of identifying and quantifying substances with either overlapping spectra and/or nonexhaustive libraries is commonly coined as "unsolvable in principle". In ^{1}H NMR spectra, ubiquitous homonuclear ^{1}H−^{1}H J-couplings generates, in particular, broad multiplets,[27] which keeps the exact elucidation of the chemical structure ambiguous.[28,29] The need of new approaches has been underlined in the review of Elyashberg et al.[30] In a similar deconvolution challenge, involving Fourier-transformed infrared spectra, we previously proposed a generalized least-squares formulation only for sparse solutions.[31] Since all substances were introduced at once, the technique was highly efficient on small libraries (23 substances initially), but was less capable on hundreds of substances or when the spectra were collected on different devices. In this paper, we propose a generalized matching pursuit decomposition, where the substances are introduced progressively, according to sequential scenarios. The choices between many scenarios or preferences are carried out using pairwise comparisons and a Schulze's rule satisfying Condorcet criterion, resolvability, Pareto optimality, reversal symmetry, and monotonicity.[32]

The paper is organized as follows. In section 2, we extend the principles of the matching pursuit decomposition in mixtures including partly correlated components. The pursuit is carried out at two scales. First, a coarse pairwise comparison restricts the search to likely substances. Second, the finest scale spans iterative and/or alternative scenarios, for which sparse solutions, minimizing the risk of false positive, are sought. Section 3 presents the ^{1}H NMR spectral database, the cases studies, and analytical methods considered in this study. Results on large-scale Monte Carlo mixture simulations and on the determination of the main additives of real plastics (synthetic and biobased ones) are presented in section 4. The robustness of the entire strategy in the presence of residues and instrumental/experimental variations are finally discussed in the last section.

2. THEORY

2.1. Background. Identifying and quantifying substances in cocktails is a generic problem coined in the signal literature with various names, such as "blind source identification", "blind deconvolution", and "sources separation problem". The chief difficulty is under-determination, which arises when the number of unknowns (spectra and concentrations) is larger than the equations describing the mixture. The undetermined system can have either no solution or an infinite number of solutions. However, by adding mild constraints, we can ensure that the original problem accepts one relevant solution or a small set of feasible ones. Possible constraints include (i) splitting the original sources between identifiable and nonidentifiable ones, and (ii) adding a general condition of sparsity (minimal number of sources).

The partitioning strategy is usually implemented via statistical projection techniques measuring the degree of independence of data such as second-order statistics[33] or high-order statistics.[34] Independent component analysis (ICA) is among the most popular.[34] As acknowledged in the recent review,[35] simple one-dimensional (1D) NMR spectra suffer a dramatic lack of resolution for substance assignment purposes. The number of overlaps increase significantly with the number of substances in the mixture, which violates the source orthogonality required by ICA.[36] Sparse component analysis (SCA),[36] which must not be confused with the condition of sparsity on the solution, was proposed a few years ago to overcome mutual overlapping by replacing a global orthogonality requirement with a local one. The method has been exemplified to the source identification up to four substances.[36−38] As original ICA, SCA works better when the same mixture of substances has been assessed with different proportions[37] or with multidimensional spectroscopic modalities.[38]

The lack of convexity (several likely solutions, intractable problems) of partitioning techniques[39] and their high sensitivity to chemical shifts errors[40] limit their use in tailored applications with minimum supervision. In this work, we preferred a reference-based substance identification involving one or several reference spectral libraries. The ability to reach a sparse solution (i.e., with a small number of likely substances) depends on how well the elements or "atoms", stored in the library, match the characteristics of the mixture. The quality of the match is maximized when the atoms are redundant and noiseless. Since the library forms a linearly dependent set, the decomposition is not unique but the entire problem can be solved with greedy or adaptive algorithms, including the matching pursuit[41] or the method of frames,[42] which have been developed in the context of signal compression and coding.

2.2. Coding NMR Signals of Reference Substances as Masks. This study prefers ^{1}H NMR spectra expressed versus chemical shifts, denoted δ, to raw free induction decay (FID) signals in the time domain. This choice offers a direct interpretation and supervision by the end-user at almost all stages of the identification and deconvolution process. End-users interested in time-domain operations should note that correlation operators applied in this study are complex multiplications in the time domain (i.e., complex conjugate of the first signal multiplied by the second). However, one adverse consequence is that spectra obtained by Fourier-transformed signals are subjected to larger biases, inherited from interactions between experimental settings (e.g., inhomogeneity of the static magnetic field, pulse phase, and flip angle errors, impurities), post-treatments (filtering, phase detection, digitization, averaging) and transition effects (Gibbs ringing). They affect the quality of spectra by causing poor baselines, frequency shifts and distortions of peaks. Only experimental white noise is propagated in a conservative manner from FID signals to spectra. All these aspects and possible misalignment of

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

theoretical or experimental spectra have been integrated within the proposed mathematical framework. Compared to most of the previous blind deconvolution methods, the number of real substances in the mixture (M) is unknown, as well as the number of substances of the library (R_M), which are present in the mixture. In this work, we assume complex mixtures and incomplete libraries, such that $R_M < M$ and M values ranging from 4 to 10. To fulfill food contact application needs, the number of substances in the reference library, R, is assumed to be large (between 50 and 10^3). Each NMR spectra in the library, denoted $\{S_k^{(\delta)}\}_{k=1,2,...,R}$, are assumed to be decomposed into a sum of T_k unitary segment functions, denoted $\{f_{k,i}^{(\delta)}\}_{i=1,...,T_k}$. Appendix S1 in ref 43 details the decomposition and coding of NMR spectra as normalized Gaussian lines (i.e., $\int_{-\infty}^{+\infty} f_{k,i}^{(\delta)}\, d\delta = 1$). Gaussian lines were preferred to Lorentzian ones, because Lorentzian tails are much more sensitive to peak broadening and baseline errors. Similar decompositions can be found in ref 44

Finally, each segment function i of the substance k is associated with a particular multiplet (resolved or not) and to a number of protons ($n_{k,i}^H$), so that $\sum_{i=1}^{T_k} n_{k,i}^H$ is equal to the total number of protons in the substance. By noticing that the intensity of the NMR signal reflects the abundance in protons, one gets

$$\frac{S_k^{(\delta)}}{\gamma C_k N_k^H} = \frac{\overline{S}_k^{(\delta)}}{\gamma N_k^H} \triangleq D_k^{(\delta)} = \frac{\sum_{i=1}^{T_k} n_{k,i}^H f_{k,i}^{(\delta)}}{N_k^H} = \sum_{i=1}^{T_k} x_{k,i}^H f_{k,i}^{(\delta)}$$
$$\text{for all } k = 1, ..., R \tag{1}$$

where $\{C_k\}_{k=1,...,R}$ is the concentration (in mol L^{-1}) in the deuterated solvent, γ is a global calibration parameter independent of the considered substance. $\{D_1^{(\delta)}, D_2^{(\delta)}, ..., D_R^{(\delta)}\}$ are a collection of R overall masks, which preserve the lines of multiplets but, which are zero outside (i.e., no baseline). As described in Appendix S1 in ref 143, they are obtained as least-squares approximants of real spectra (measured or predicted ones). By analogy with the language, they represent the "pronunciation of words" (the masks) but not the exact "sounds" (the spectra). As a result, the "words" can be recognized in a "speech" (mixture) even when they are pronounced by different speakers. Finally, as in real-life speeches, some "words" can be missing from the "dictionary" of the listener (library) and are consequently not understood, but without hindering the identification of known ones.

2.3. Principles of the Matching Pursuit Decomposition. In the context of NMR spectroscopy, the spectrum of any mixture is written as a weighted sum of molar spectra $\overline{S}_k^{(\delta)}$, which can be factorized into two groups, on the basis of whether the substances are included in library or not. By denoting the two groups s_{in} and s_{out}, respectively, the decomposition reads

$$S_{\text{mixture}}^{(\delta)} = \sum_{k=s_{in}} C_k \overline{S}_k^{(\delta)} + \sum_{k=s_{out}} C_k \overline{S}_k^{(\delta)} + B^{(\delta)} + \epsilon^{(\delta)}$$
$$= \sum_{k=s_{in}} C_k \overline{S}_k^{(\delta)} + S_{\text{irreductible residue}}^{(\delta)} + \epsilon^{(\delta)} \tag{2}$$

where $B^{(\delta)}$ is a "smooth" baseline, whose height is smaller than the lines of interest and $\epsilon^{(\delta)}$ is random white noise, respectively.

The pursuit decomposition algorithm replaces the normalization by a reference peak[45,46] by several matches with theoretical lines. The simplest situation occurs when the identity of substances in the mixture and included in s_{in} is known. The concentrations $\{C_k\}_{k\in s_{in}}$ are hence determined successively by comparing $S_{\text{mixture}}^{(\delta)}$ with the masks of s_{in} stored in the library. When all masks $\{D_k^{(\delta)}\}_{k=1,...,R}$ constitute an orthogonal basis, their inner products are zero:

$$\langle D_u^{(\delta)}, D_v^{(\delta)} \rangle = 0 \qquad \text{for all } u = 1, ..., R;\ v = 1, ..., R;\ u \neq v \tag{3}$$

and $\{C_k\}_{k\in s_{in}}$ are inferred directly as projection coefficients:

$$\hat{C}_k = \frac{1}{\gamma N_k^H} \langle S_{\text{mixture}}^{(\delta)}, D_k^{(\delta)} \rangle \qquad \text{for all } k \in S_{in} \tag{4}$$

where $\langle f^{(\delta)}, g^{(\delta)} \rangle$ is the Euclidian scalar product of two discretized NMR spectra $\langle (f(\delta_1), f(\delta_2), ..., f(\delta_n)), (g(\delta_1), g(\delta_2), ..., g(\delta_n)) \rangle = \sum_{i=1}^{n} f(\delta_i) g(\delta_i)$. It is important to notice that this product represents also the cross-correlation operator at zero lag.

When the substances of s_{in}, which are present in the mixture, are unidentified, the previous method can be generalized by applying iteratively eq 4 to the substance, which exhibits the highest similitude with the remaining residue. In the standard matching pursuit algorithm, the remaining spectral residue replacing $S_{\text{mixture}}^{(\delta)}$ in eq 4 is given by

$$S_{\text{residue}}^{(\delta),n} = S_{\text{residue}}^{(\delta),n-1} - \langle S_{\text{residue}}^{(\delta),n-1}, D_{k_{\text{best}}}^{(\delta)} \rangle D_{k_{\text{best}}}^{(\delta)} \tag{5}$$

with $S_{\text{residue}}^{(\delta),0} = S_{\text{mixture}}^{(\delta)}$ and k_{best} the index of the "best" matching substance in the library. The likeliest substance is the substance k_{best}, which maximizes $\langle S_{\text{residue}}^{(\delta),n-1}, D_{k_{\text{best}}}^{(\delta)} \rangle$.

2.4. Projecting onto Nonorthogonal Segment Functions. Equation 5 enables one to recognize a pattern even if its shape has been slightly modified. When some masks in the library are partly correlated together, the pursuit method cannot, however, be applied explicitly and a linear system of equations needs to be resolved instead. Using the u^{th} mask in the library as test 1 and the white noise property $\langle \epsilon^{(\delta)}, D_u^{(\delta)} \rangle \approx 0$, eq 2 becomes

$$\langle S_{\text{mixture}}^{(\delta)}, D_u^{(\delta)} \rangle = \sum_{k=1,...,R} C_k \langle \overline{S}_k^{(\delta)}, D_u^{(\delta)} \rangle + \langle S_{\text{irreductible residue}}^{(\delta)}, D_u^{(\delta)} \rangle$$
$$\text{for } u = 1, 2, ..., R \tag{6}$$

under the constraints $\{C_k\}_{k=1,...,R} \geq 0$.

The assembled system contains as much linear equations as unknowns $\{C_k\}_{k=1,...,R}$ but it still remains a strongly ill-posed problem. The first complication arises from the unknown term $\langle S_{\text{irreductible residue}}^{(\delta)}, D_u^{(\delta)} \rangle$, which enforces a resolution of the system described by eq 6 only in a least-squares sense. The second complication originates from the size R of the library itself. Only sparse solutions maximizing the number of concentrations equal to zero ($C_k = 0$) are very likely.

A reliable regularization strategy consists of spanning each mask into its segment functions via eq 1 and introducing redundant information in the original least-squares problem. For the uth mask, the distributivity of the inner product replaces eq 6 via T_u relationships:

$$\sum_{k=1,...,R} \langle \overline{S}_k^{(\delta)}, f_{u,i}^{(\delta)} \rangle C_k = \langle S_{\text{mixture}}^{(\delta)}, f_{u,i}^{(\delta)} \rangle - \langle S_{\text{irreductible residue}}^{(\delta)}, f_{u,i}^{(\delta)} \rangle$$
$$\text{for } u = 1, 2, ..., R \text{ and } i = 1, ..., T_u \tag{7}$$

which are recast in matrix notations as

$$\mathbf{S_{RR}C} = \mathbf{S_{mixture}} - \boldsymbol{\sigma}_{\text{residue}} \tag{8}$$

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

where $\mathbf{S_{RR}} = (\langle \overline{S}_k^{(\delta)}, f_{u,i}^{(\delta)} \rangle)_{u=1,...R;\ i=1,...,T_u;\ k=1,...,R}$ is a $(\sum_{u=1,...,R} T_u) \times R$ matrix (the subscript RR is doubled to distinguish matrices from vectors), $\mathbf{C} = (\hat{C}_k)_{k=1,...,R}$ is a $R \times 1$ vector, $\mathbf{S_{mixture}} = (\langle S_{\text{mixture}}^{(\delta)}, f_{u,i}^{(\delta)} \rangle)_{u=1,...,R,\ i=1,...,T_u}$ and $\boldsymbol{\sigma}_{\mathbf{residue}} = (\langle S_{\text{irreductible residue}}^{(\delta)}, f_{u,i}^{(\delta)} \rangle)_{u=1,...,R,\ i=1,...,T_u}$ are $(\sum_{u=1,...,R} T_u) \times 1$ vectors.

2.5. Generalized Matching Pursuit Decomposition. Overdetermination of the formulation presented in eq 8 is similar to a cross-correlation matching procedure whose objective is minimization of the contribution of residues, $\boldsymbol{\sigma}_{\text{residue}}$, in the space spanned by the segment functions stored in the library. When substances contributing to residues are very dissimilar, the components of $\boldsymbol{\sigma}_{\text{residue}}$ are alike to be close to normal variates with zero mean. Conversely, the presence within residues of highly common functions (e.g., CH_2 groups) or homologous substances (e.g., oligomeric additives or including similar chemical functions) may increase the risk of generating false positives (substances wrongly identified). The risk can be mitigated by weighting segment functions according to their ability to discriminate substances belonging to the same family and by introducing some parsimony principles. Previous considerations lead to the following non-negative least-squares problem:

$$\min_C \frac{1}{2}(\mathbf{W}\sigma_{\text{residue}})^{\mathrm{T}}\mathbf{W}\sigma_{\text{residue}} = \min_C \frac{1}{2}\|\mathbf{WS_{RR}C} - \mathbf{WS}_{\text{mixture}}\|^2 \quad \text{subject to } C \geq 0 \tag{9}$$

where $\|\cdot\|$ denotes the L_2 norm, $\mathbf{W}$ is a $(\sum_{u=1,...,R} T_u) \times (\sum_{u=1,...,R} T_u)$ diagonal matrix coding for the weight of each segment function. One reasonable choice is to choose weights, which are proportional to their reciprocal occurrence in the library. To keep a neutral identification (all substances are equiprobable), the sums of weights related to each substance $u = 1, ..., R$, $(\sum_{i=1,...,T_u} w_{u,t})$ must finally be forced to be equal. A similar approach, but based on more restrictive assumptions, has been suggested independently for Raman spectra.[39] The authors assume that R is not too large, that the upper bounds of their concentrations are known and that spectral lines are not subjected to any squeeze or shift. Conversely to the approach followed by authors, the constraint on the non-negativity of residues is not considered, because it would affect the normality of errors.

The non-negativity constraint can be enforced via various mathematical methods. They are shortly reviewed with notations of eq 9. The most-used method is derived from the seminal book of Lawson and Hanson[47] and subsequently implemented in Matlab/Octave as lsqnonneg. The basic idea is to minimize $J = {}^1/_2(\mathbf{W}\boldsymbol{\sigma}_{\mathbf{residue}})^{\mathrm{T}}\mathbf{W}\boldsymbol{\sigma}_{\mathbf{residue}}$ iteratively while finding a solution to equation $\mathbf{WS_{RR}C} = \mathbf{WS_{mixture}}$. The procedure is slightly similar to the standard original matching decomposition. The solution is initialized to $\mathbf{C} = 0$ (no substance from the library is introduced) and at each iteration the substance, which contributes the most to the gradient of J (equal to $\mathbf{S}_{\mathbf{RR}}^{\mathrm{T}}\mathbf{W}^2[\mathbf{S_{mixture}} - \mathbf{S_{RR}C}]$) is appended to the list of active substances. At each iteration, the system is solved in a least-squares sense by considering only active substances, while a null concentration is assigned to others. The algorithm stops as soon as J ceases decreasing. As reported by Chen and Plemmons,[48] the convergence of this method is usually acceptable and yields solutions varying from "fairly good" (optimal) to very likely (suboptimal) solutions. For very large ill-conditioned systems (e.g., thousands substances), speed gain can be achieved either by precomputing parts of the pseudo-inverse of $\mathbf{S}_{\mathbf{RR}}^{\mathrm{T}}\mathbf{S_{RR}}$[49,50] or by incorporating more than one substance at once: via a gradient projection step,[51] a sequential coordinate-wise algorithm,[52] an interior point method[53] or a block pivoting method.[54]

2.6. Optimized Cross-Correlation Matching for Both Substance Identification and Quantification. Strategies transforming some inequality constraints (non-negative concentrations) into equality ones (zero concentrations) minimize on purpose the risk of false positives. Finding a good set of substances to initiate the decomposition (in other words, setting equality constraints to all others) remains a challenge. In previous methods, the choice is oriented by local components of $(\partial J/\partial C_k)_{k=1,...,R}$, which improve the most the global regression coefficient. However, this strategy suffers adverse risks of overfitting and zigzagging: among two or more possible substances, only one will be picked without exploring alternatives with a significant sensitivity to residual biases and noise. In the context of blind clustering of ^{1}H NMR spectra, Sun and Xin[44] have partly solved this issue by seeking at each substep the dominant spectrum interval (by solving a non-negative L_1 problem with either a projected gradient descent or with linearized Bregman iterations) maximizing the separation of sources. The reported method is essentially qualitative, but it offers a strong robustness regarding the width and overlapping rate of the peaks. In our work, we followed similar ideas but with a different method. Robustness was improved by replacing, in eqs 7 and 9, the correlation operator $\langle f^{(\delta)}, g^{(\delta)} \rangle$ by the maximum of the cross-correlation, which is insensitive to translation within the range $-d_{\max}$ and $+d_{\max}$:

$$\langle f^{(\delta)}, g^{(\delta)} \rangle_{\max} = \max_{-d_{\max} \leq d \leq +d_{\max}} [(f^{(\delta)} \bigstar g^{(\delta)})(d)] \tag{10}$$

with

$$(f^{(\delta)} \bigstar g^{(\delta)})(d) = (f^{(\delta)} \otimes g^{(-\delta)})(d) = \int_{-\infty}^{+\infty} f^{(\delta)} g^{(\delta+d)}\, \mathrm{d}\delta$$

where $\otimes$ denotes convolution. The new operator (eq 10) can be calculated efficiently via fast Fourier transform, and the $(\sum_{u=1,...,R} T_u)^2$ values of $\mathbf{S_{RR}}$ can be precomputed and stored for future uses. Only the $\sum_{u=1,...,R} T_u$ values of $\mathbf{S}_{\text{mixture}}$ need to be recalculated for each new sample. In practice, $d_{\max}$ is kept small and chosen between 0.02 and 0.15 ppm for experimental spectra and up to 0.3 ppm for theoretical spectra.

The rapid increases of the condition number of the matrix $\mathbf{S_{RR}}$ with the number of substances in s_{in} prevents the direct minimization of eq 9 in the presence of many correlated substances. We restrict the blind deconvolution to the $R_{\text{choice}} \approx 10 \ll R$ likeliest substances (i.e., $\mathbf{S_{RR}}$ is accordingly replaced by its shortest version $\mathbf{S}_{\mathbf{R}_{\text{choice}}\mathbf{R}_{\text{choice}}}$) by choosing an introduction criterion, similar to eq 4. The global regression coefficient $\{b_{\text{mixture},k}\}_{k=1,...,R}$ between multiplet functions and $S_{\text{mixture}}^{(\delta)}$ offers a natural choice:

$$\{b_{\text{mixture},k}\}_{\text{likely } k\text{th substance}} = \sum_{i=1,...,T_k} w_{k,i} \frac{\langle \mu^{(\delta)} S_{\text{mixture}}^{(\delta)}, f_{k,i}^{(\delta)} \rangle_{\max} - \overline{\mu^{(\delta)} S_{\text{mixture}}^{(\delta)}}\ \overline{f_{k,i}^{(\delta)}}}{\langle f_{k,i}^{(\delta)}, f_{k,i}^{(\delta)} \rangle_{\max} - \overline{f_{k,i}^{(\delta)}}^2} \geq b_{\text{threshold}} \tag{11}$$

where $\overline{X}$ is the arithmetic average of X; $\mu^{(\delta)} = \sum_{k=1,...,R} D_k^{(\delta)}$ is a global mask restricting the comparison on regions matching the

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

segment functions in the library; $b_{\text{threshold}}$ is a threshold constant controlling the tolerance to noise.

The real number of substances in the library present in the mixture (R_M) can be estimated heuristically by comparing the regression coefficient involving the entire database with the one associated with the likeliest $k_{\max}$th substance:

$$R_M \approx \frac{\sqrt{\dfrac{\langle \mu^{(\delta)} S^{(\delta)}_{\text{mixture}}, \mu^{(\delta)} S^{(\delta)}_{\text{mixture}} \rangle_{\max} - \overline{\mu^{(\delta)} S^{(\delta)}_{\text{mixture}}}^{\,2}}{\langle f^{(\delta)}_{k,i}, f^{(\delta)}_{k,i} \rangle_{\max} - \overline{f^{(\delta)}_{k,i}}^{\,2}}}}{b_{\text{mixture},k_{\max}}} \tag{12}$$

Although the proposed procedure selects substances based on multiple spectral similarities, it also has a tendency to select substances with collinear spectra. Ambiguous insertions of substances can be assessed by calculating pairwise correlation coefficients, $\{\rho_{k_1,k_2}\}_{k_1,k_2=1,\ldots,R}$, in the same way as eq 11:

$$\rho_{k_1,k_2} = \sum_{i=1,\ldots,T_{k_2}} w_{k_2,i} \times \frac{\langle \mu^{(\delta)} \bar{S}^{(\delta)}_{k_1}, f^{(\delta)}_{k_2,i} \rangle_{\max} - \overline{\mu^{(\delta)} \bar{S}^{(\delta)}_{k_1}}\ \overline{f^{(\delta)}_{k_2,i}}}{\sqrt{(\langle \mu^{(\delta)} \bar{S}^{(\delta)}_{k_1}, \mu^{(\delta)} \bar{S}^{(\delta)}_{k_1} \rangle_{\max} - \overline{\mu^{(\delta)} \bar{S}^{(\delta)}_{k_1}}^{\,2})(\langle f^{(\delta)}_{k_2,i}, f^{(\delta)}_{k_2,i} \rangle_{\max} - \overline{f^{(\delta)}_{k_2,i}}^{\,2})}} \tag{13}$$

and by clustering the substances according to the distance matrix

$$\left\{ \frac{1}{\sqrt{\rho_{k_1k_2}\rho_{k_2k_1}} + \epsilon} - 1 \right\}_{k_1,k_2=1,\ldots,R}$$

with ϵ being a small number.

Equations 11−13 provide a set of sufficient constraints to build a list of preferences not only among the R substances of the library but also among the R_{choice} substances. In this work, we choose to build sparse solutions based on the concept of the Condorcet winner. In a particular ranked ballot scenario, a substance of class Γ and rank r "wins" its "seats" in the solution, if it provides the highest regression coefficient of its class and if the next substance in class Γ is ranked $r + 2$ or greater. When a substance of class Γ ranked $r + 1$ exists, the ranking is not considered to be transitive (preferences could be reverse); an alternative scenario considering the second (or third, etc.) substance of class Γ possibly present is added. These principles are illustrated for a set of seven substances (sorted in decreasing order of preference): A1 (the most similar), B1, B2, C1, C2, C3, and D1 (the least similar). (In this nomenclature, the letter represetns the class and the digit represents the position in their class.) All 64 possible preference lists (similar to votes) are plotted as a hierarchical tree in Figure 1. It reads as follows:

- 1 preference list of length 1 (A1),
- 2 of length 2 ({A1 B1} and {A1 B2}),
- 17 of length 3,
- 21 of length 4,
- 15 of length 5,
- 6 of length 6, and
- 1 of length 7 ({A1 B1 B2 C1 C2 C3 D1}).

At this stage, it is important to note that selecting a branch based on Borda counts (i.e., number of substances ranked lower) or an equivalent strategy suffers a lack of axiomatic properties,[55] including

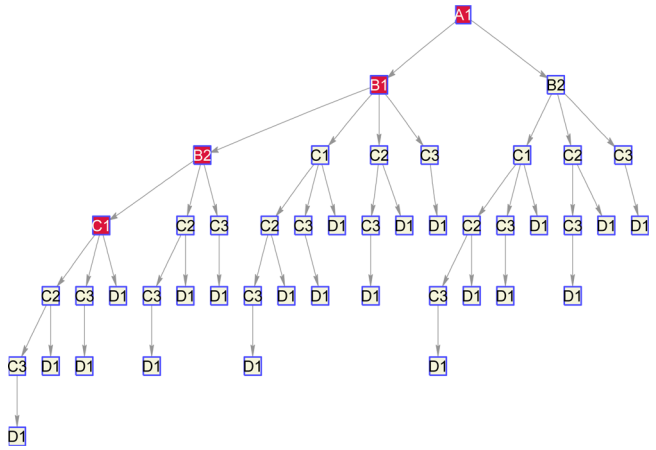


Figure 1. Example of insertion lists for seven substances coded as A1 > B1 ≥ B2 > C1 ≥ C2 ≥ C3 > D1, where the symbol ">" means "preferred to" and the symbol "≥" means "preferred or equivalent to". In this nomenclature, letters A−D denote the chemical class of the substance and the following digit represents its local rank in the class. The optimal scenario of length 4 is depicted with colored nodes.

- no guarantee to pick the preferred substance (Condorcet criterion),
- no other substance combination outperforms the current selection (Pareto efficiency),
- the selection is immune to irrelevant alternatives (monotonicity criterion), and
- the score of a selection must not be affected by the addition of similar substances (independence of clones criterion).

Both the ranked pairs rule[56] and the Schulze's one[32] satisfy previous rules. The popular Schulze voting method has been preferred, because it accommodates unranked substances/candidates well and many implementations have been proposed by the open-source community. In the example shown in Figure 1, it gives, consistently, A1 B1 B2 C1 as the preferred path of length 4. The non-negative solutions of eq 9 would ultimately determine whether B1 or B2 or both are present in the mixture.

3. MATERIALS AND METHODS

3.1. ^{1}H NMR Spectral Database. The database used for this study comprised experimental ^{1}NMR spectra (including repetitions and spectra at different concentrations) of 52 substances, as reported in Table 1 and detailed in Table S1 in the Supporting Information.[43] All the main technological classes of substances were represented: antioxidants, light stabilizers, plasticizers, photoinitiators, biocides, monomers, slip additives. The database was also augmented with the theoretical spectra of most common substances found in plastic materials to show the capability of the method to work with misaligned spectra. Theoretical spectra were calculated using ChemNMR Pro software from ChemBioDraw Ultra 14 Suite,[57] assuming deuterated chloroform solvent as in our experiments. All measured and theoretical spectra of pure substances were decomposed as Gaussian lines, according to eqs (S1)−(S3) of Appendix S1 in the Supporting Information[43] and regenerated as spectra via eq 1. Encoding multiplets via a set of elementary mathematical functions offered high compression ratios of NMR spectra and robust baseline suppression. Corresponding

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

Table 1. Detailed Information of Substances in the Spectral Database in Our Study[a]

code	common name (CAS number)	technological class[b]	chemical class[c]	code	common name (CAS number)	technological class[b]	chemical class[c]
M01	BHT (128-37-0)	A	1	M02	Irganox 1010 (6683-19-8)	A	1
M03	Irganox 1076 (2082-79-3)	A	1	M04	Irganox 1330 (1709-70-2)	A	1
M05	Irganox 1520 (110553-27-0)	A	1	M06	Irganox 245 (36443-68-2)	A	1
M07	Irganox 3052 (61167-58-6)	A	1	M08	Irganox 3114 (27676-62-6)	A	1
M09	Irgafos 168 (31570-04-4)	A	2	M10	Irganox 1035 (41484-35-9)	A	1
M11	Irganox PS 800 (123-28-4)	A	1	M12	Irganox PS 802 (693-36-7)	A	1
M13	BIT (2634-33-5)	B	(−)	M14	CMIT (26172-55-4)	B	(−)
M15	MIT (2682-20-4)	B	(−)	M16	DPGME (34590-94-8)	co; s	(−)
M17	MBOCA (101-14-4)	cu	(−)	M18	TEP (122-52-1)	dh	(−)
M19	stearic acid (57-11-4)	L	(−)	M20	Bisphenol A (80-05-7)	m	(−)
M21	Caprolactam (105-60-2)	m	(−)	M22	HMDA (124-09-4)	m	(−)
M23	Resorcinol (108-46-3)	m	(−)	M24	4BBP (2128-93-0)	ph	(−)
M25	4HBP (1137-42-4)	ph	(−)	M26	ITX (75081-21-9)	ph	(−)
M27	4MBP (134-84-9)	ph	(−)	M28	DEHA (103-23-1)	P	3
M29	butylbenzoate (136-60-7)	P	4	M30	DPGDB (27138-31-4)	P	4
M31	tributylacetylcitrate (77-90-7)	P	5	M32	triethylcitrate (77-93-0)	P	5
M33	Santicizer 141 (1241-94-7)	P	6	M34	BBP (85-68-7)	P	7
M35	DBP (84-74-2)	P	7	M36	DEHP (117-81-7)	P	7
M37	dibutylseba-cate (109-43-3)	P	8	M38	PE (9002-88-4)	M	(−)
M39	PP (9003-07-0)	M	(−)	M40	PS (9003-70-7)	M	(−)
M41	diphenylthio-urea (102-08-9)	st	(−)	M42	erucamide (112-84-5)	sl	(−)
M43	oleamide (301-02-0)	sl	(−)	M44	acetophenone (98-86-2)	S	(−)
M45	diethylenegly col (111-46-6)	S	(−)	M46	Chimassorb 81 (1843-05-6)	UV	9
M47	Tinuvin 326 (3896-11-5)	UV	10	M48	Tinuvin P (2440-22-4)	UV	10
M49	Chimassorb 944 (71878-19-8)	UV	11	M50	Tinuvin 622 (65447-77-0)	UV	11
M51	Tinuvin 770 (52829-07-9)	UV	11	M52	benzophenone (119-61-9)	UV; ph	9

[a]Substances are ordered by their technological class. Technological chemical classes were derived from refs 58−62. [b]Legend for technological class: A, antioxidant; B, biocide; co, coalescing agent; cu, curing agent; dh, hyperoxide decomposer, L, lubricant; m, monomer; M, polymer; ph, photoinitiator; P, plasticizer; sl, slip additive; st, PVC stabilizer; S, solvent; and UV, UV stabilizer. [c]Legend for chemical class: 1, phenol; 2, phosphite-phophonite; 3, adipate; 4, benzoate; 5, citrate; 6, phosphate, 7; phthalate; 8, sebacate; 9, benzophenone; 10, benzotriazole; and 11, hindered amine. Pure substances were provided from Sigma−Aldrich (France), TCI Europe N.V. (Belgium), Alfa Aesar (USA), and VWR International (France).

coefficients were stored into a common database in XML (eXtensible Markup Language).

3.2. Reference Materials. Five processed materials with known formulations, samples denoted as S3−S7, were used to test the entire identification and quantification methodology. Materials and formulations subsequently inferred after polymer extraction are detailed in Table 2. Considered additives were chosen to be representative of common applications of considered plastics, including UV stabilizers, slip agents, plasticizers, and photoinitiators.

Two high-density polyethylene (HDPE) materials, identified in this study as samples S3 and S4, were formulated by twin-screw extrusion at an industrial scale. The third material in polypropylene (PP), identified as sample S5, was formulated in our laboratory with six additives by immersion. Virgin PP material was cut into strips (60 mm × 3 mm in size) and immersed into a dichloromethane (Sigma−Aldrich, France) solution at 100 g L^{-1} for 3 days at 60 °C in pressurized vials. Formulated strips were subsequently rinsed and dried at room temperature for ∼30 min. All formulated materials were stored at ambient temperature in hermetically closed vials before use. Samples S6 and S7 were prepared from polylactide (reference PLA - PLI 003, Natureplast, Ifs, France) as a prototype of biosourced polymers. They were processed on a semi-industrial scale, as detailed in ref 63. Both materials contained four additives; in particular, sample S6 also included 1-hexadecanol, which was not present in the assembled library.

Extraction by maceration in a stirred solvent was used to determine the concentrations in reference materials. One gram of materials cut into small pieces was placed in contact with 20 mL of solvent at room temperature. Chloroform (VWR International, France) was used for polyolefin materials (samples S3−S5) and mixture of methanol/chloroform (80:20 v:v) was preferred for bio-based polymers (samples S6 and S7). Each extraction flask was placed on an orbital stirrer with 200 rounds per minute for at least 48 h. Turbid extracts were filtered on glass wool, using a polytetrafluoroethylene (PTFE) syringe with a pore size of 0.45 μm (Sigma−Aldrich, France) prior analyses. After extraction, solvent was evaporated. Dry samples were stored at −18 °C for a maximum of 7 days.

Reference concentrations of extracts S3−S7 were determined using both high-performance liquid chromatography (HPLC) and gas chromatography (GC). The HPLC system (Model 717plus Autosampler, Waters, USA) was composed of a thermostated column (model Xbridge Phenyl, 150 mm × 3.0 mm × 5 μm; Model 600 Controller, Waters, USA) and a photodiode array detector (Model 2996 PDA Detector, Waters, USA). The eluent gradient system consisted of 0.5 g L^{-1} ammonium acetate (eluent A), acetonitrile (eluent B), and tetrahydrofuran (eluent C). The composition of applied gradient was as follows:

85% A, 5% B, and 10% C) at 0 min

5% A, 85% B, and 10% C at 40 min

85% A, 5% B, and 10% C at 60 min

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

Table 2. Formulated Materials (Samples S1−S7) Used in Challenge Tests[a]

	Numerical Examples (Relative Composition)		Processed Materials (Composition, in g kg^{-1})				
	S1	S2	S3	S4	S5	S6	S7
difficulty	−	++	+	+++	+	+++	++
overlapping level	−	++	−	+++	++	+++	++
polymer	N/A	N/A	HDPE	HDPE	PP	PLA	PLA
extraction solvent	not required			$CHCl_3$		$CH_3OH/CHCl_3$ (80:20,-v/v)	
post-treatments	not required		drying and dissolution of the extract in $CDCl_3$				
Irganox 1010 (M02)	0	0	0	0	8.5	0	0.1
Irganox 1076 (M03)	0	random proportions between 0.1 and 0.9	0	0	0.9	0.9	0
Irganox 3114 (M08)	0	0	0	0	0	0	0.3
Irgafos 168 (M09)	0	0	0	>0	0	0	0
Irganox 1035 (M10)	0	random proportions between 0.1 and 0.9	0	0	0	0	0.4
Irganox PS802 (M12)	0	0	0	0	0	0.4	0
CMIT (M14)	random proportions between 0.1 and 0.9	0	0	0	0	0	0
stearic acid (M19)	0	0	0	0	0	0	0.4
Bisphenol A (M20)	random proportions between 0.1 and 0.9	0	0	0	0	0	0
DEHA (M27)	0	0	0	0	7.9	0	0
triethylcitrate (M32)	random proportions between 0.1 and 0.9	0	0	0	0	0	0
erucamide (M42)	0	random proportions between 0.1 and 0.9	9.5	0	1.3	0	0
Chimassorb 81 (M46)	0	0	0	0	0	0.5	0
Tinuvin 326 (M47)	random proportions between 0.1 and 0.9	random proportions between 0.1 and 0.9	0	0	12.2	0	0
Chimassorb 944 (M49)	0	0	0	8.5	0	0	0
Tinuvin 622 (M50)	0	0	0	10.5	0	0	0
benzophenone (M52)	0	0	0	0	7.7	0	0
1-hexadecanol	0	0	0	0	0	1	0

[a]Formulations of samples S3−S7 were assessed after polymer extraction.

Table 3. Minimum Correlation Coefficients To Confuse Substances of Mixtures S1 and S2 (see Table 2) with Other Substances in the Library (Based on Correlation Matrix Depicted in Figure S2 in the Supporting Information[43]

mixture	M03	M10	M14	M20	M32	M42	M47
S2	×	×				×	×
S1			×	×	×		×

	Critical Pairwise Correlation Coefficients						
number of confused substances	M03	M10	M14	M20	M32	M42	M47
0	0.967[a]	0.993[a]	0.985[a]	0.976[a]	0.976[a]	0.977[a]	0.983[a]
1	0.627	0.452	0.651	0.467	0.58	0.891[a]	0.551
2	0.526	0.357	0.494	0.415	0.5	0.562	0.456
3	0.503	0.309	0.473	0.401	0.462	0.456	0.438
4	0.387	0.281	0.466	0.398	0.452	0.437	0.414

[a]This correlation coefficient is larger than the threshold retained in this study (0.75).

GC conditions consisted of a column (Model VF-5MS, 30 m × 0.25 μm × 0.25 μm, Varian FactorFour, USA) at 275 °C mounted on a Model 431-GC system (Varian, USA) coupled with an ion-trap mass spectrometer (Model 220-MS, Varian, USA).

3.3. ^{1}H NMR Measurements. Pure additives and dry extracts of reference materials were diluted in 1 mL of 99.8% deuterated chloroform (Sigma−Aldrich, France). Their ^{1}H NMR spectra were acquired on a Bruker 360 MHz NMR spectrometer at 298 K, between −1 ppm and 14 ppm, with a relaxation time (D_1) of 25 s and a resolution of 2^{15} points. Spectra were accumulated over 32 scans (1 sample/30 min). Proton NMR spectra were obtained from Fourier transform of free induction decay (FID) signals. The singlet of tetramethylsilane (TMS) was chosen as reference frequency. Spectra were subsequently phase-shifted to remove the dispersion component. These treatments were carried out with ACD/NMR software (version 12.01, ACDLabs, USA).

4. RESULTS AND DISCUSSION

4.1. Spectral Decomposition of NMR Signals. Proton spectra of reference substances were decomposed into a sum of segments or multiplets (see eq 1), whose positions, shapes, and spans were stored within the library for pattern identification (279 multiplets are stored in this study). In our generalized matching pursuit decomposition (see eqs 7 and 10), a multiplet of multiplicity L is identified by a cross-correlation function presenting exactly $2L - 1$ peaks. A match can be obtained by translating each multiplet independently within acceptable ranges. Figure S1 in the Supporting Information[43] presents common multiplets and their autocorrelation coefficients corresponding to theoretical matches.

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

Chemical groups shared by several additives or oligomers complicate the identification and quantification of substances significantly. Without proper weighting, the pursuit algorithm would have prioritized substances, which present the highest proton abundance or generic patterns such as aliphatic segments. The importance of the signal of labile protons (mainly singlets) was also lessened, since they do not provide reliable signatures in mixtures. For each substance k, previous biases were mitigated by choosing relative weights, $\{w_{k,i}\}_{i=1,...,T_k}$, which followed the Pareto principle: the scarcity of each multiplet in the assembled library created its value. The corresponding pairwise correlation matrix, calculated from eq 13, is given in Figure S2 in the Supporting Information.[43] Its main interest is that each pair compares a reference spectrum reconstructed from eq 1 with an experimental one, as if it was a real mixture. Diagonal values were kept close to unity, ranging from 0.909 to 0.998 with a median value of 0.972, and confirmed the quality of the fit of each spectrum by a sum of Gaussian lines. The risk of confusion was estimated by off-diagonal terms, which exhibited a quite good sparsity (80% of values are lower than 0.346). The strongest similarities appeared for substances sharing either common chemical structures or technological functions (e.g., Irganox PS800 and Irganox PS802, erucamide and oleamide, etc.).

The number of substances, which could be confused from our decomposition of NMR spectra, is illustrated in Table 3 on substances of mixtures S1 and S2 (see Table 2). The risk to mistake one substance by another was assessed by calculating the critical threshold to get one or more substances with spectra similar to any substance in the mixture. The risk was high when the correlation threshold was low. The risk to confuse at least one substance occurred for correlation coefficients equal to or lower than 0.891. In other words, a substance of S1 or S2 could be identified undoubtedly only if it exhibited a correlation coefficient with the mixture greater than 0.891.

4.2. Classification of Substances Versus Projection Pursuit Method. *4.2.1. Classification Tree.* An agglomerative hierarchical classification was set by assembling a symmetric distance matrix based from pairwise correlation coefficients as $1/(\rho_{k_1k_2}\rho_{k_2k_1})^{1/2} - 1$. Generated distances were scale invariant (i.e., independent of concentration ranges and conditions of measurements), insensitive to baseline shifts, and showed good separability features, thanks to the filtering properties of cross-correlation functions. Clustering was based on average distances and back-mapped to the same scale of correlation coefficients to infer critical correlation coefficients for each substance family. The full spanned tree (including 62 classes) and dendrogram with optimal pruning at a threshold of ρ_{class} = 0.75 (including 18 classes) are presented in Figure 2. Repeated spectra of the same substance, denoted as R1−R5, enabled us to validate the reliability of the overall classification. In the perspective of pursuit decomposition, the substance or its family was much easier to identify, because it separated from others at a lower threshold. Other than the pairs Irganox PS800 and PS802, and oleamide and erucamide, which were only separated at high thresholds almost within experimental errors, all other substances were identifiable, with thresholds below 0.85.

The optimum threshold was chosen as a tradeoff between the number and size of classes and the sensitivity of the pursuit algorithm. Choosing a threshold that is too high would lead to

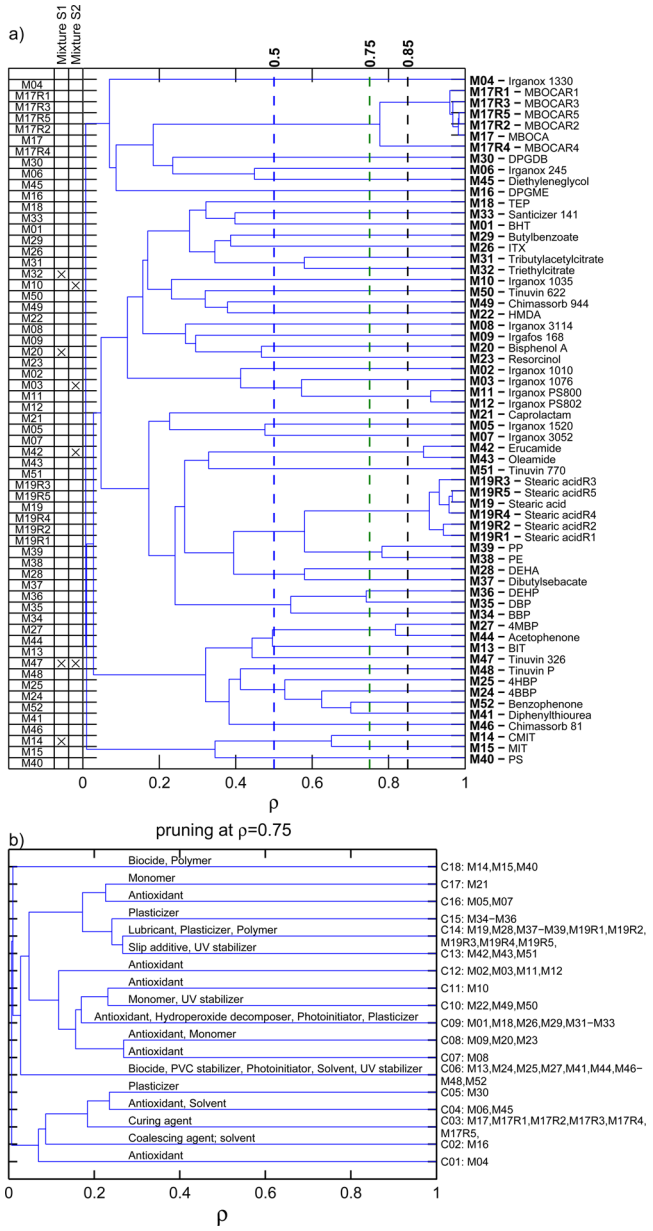


Figure 2. Classification trees of ^{1}H NMR spectra based on their pairwise correlation coefficients: (a) fully spanned tree (62 independent classes) and (b) optimally pruned tree above a correlation coefficient of 0.75 (18 classes).

no valuable matches, whereas a value that is too low might create a significant number of false positives. In order to determine the optimum, the pairwise correlation coefficients $\{(\rho_{k_1k_2}\rho_{k_2k_1})^{1/2}\}_{k_1 \neq k_2}$ within a same class were considered as a random variable. Their distributions were studied according to the threshold value (denoted as ρ_{class}). Figures 3a and 3b plot the distribution of the number of classes and of the corresponding number of substances per class at prescribed ρ_{class} values, respectively. High percentiles exhibited a "L" shape with a corner located when trees included 15−20 classes (see point C in Figure 3a). Above this number of classes, the separation was excellent or fair for the less-populated classes and poor below this number of classes. This critical number of classes corresponded to an average correlation coefficient of 0.75 with classes consisting of one single substance to 11

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

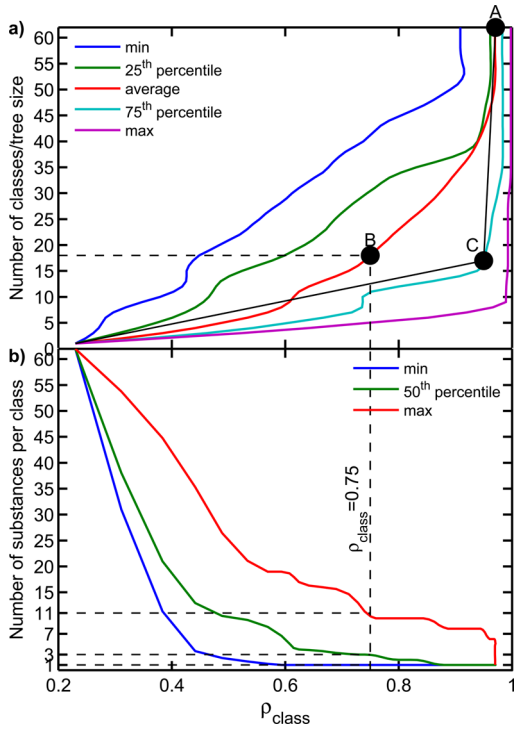


Figure 3. Distributions of (a) pairwise correlation coefficients according to the size of the spanned tree and (b) the number of substances per class according to the applied pruning cutoff. Points denoted "A" (best fit), "B" (average expectation), and "C" (corner) are vertices used in the text to set the tradeoff between identifiability and separability.

substances per class. On average, each class consisted of three substances. The details of the tree are depicted in Figure 3b.

4.2.2. Matching Pursuit Ranking. Pairwise correlation coefficients between theoretical spectra provided only an indirect assessment of the separability of substances. The performances in complex mixtures were tested numerically on mixtures S1 and S2, which consist of four substances, as reported in Table 2, and in proportions that randomly vary between 0.1 and 0.9. The results, averaged over 10^3 random mixtures, are summarized in Figures 4 and 5 as Pareto charts. Substances are plotted in increasing order, respective to their similitude with the spectrum of the full mixture. In the general case, the condition of orthogonality necessary for orthogonal matching pursuit and the condition of uncorrelatedness are not equivalent. They match only if the signal outside multiplet lines has a zero average. The presence of nonzero segments or baselines outside the coded region of each substance biased artificially correlation coefficients by adding a contribution to noncoding regions. This effect was analyzed by repeating the entire process with mixtures combining only Gaussian lines instead of real spectra. The lost details and baseline improvements are magnified in Figures 4a and 5a, respectively.

Without the property of orthogonality, the best matching substance in the remaining list cannot be determined without deconvolution and the comparison must be processed in bulk. The list of substances likely to be present in the mixture must include the best matching one as well as the following ones until a sharp drop in pairwise correlation coefficients is observed or until a correlation threshold is crossed (e.g., 0.75). Substances belonging to the same cluster are more likely to appear consecutively and, without complementary information, there is no argument to prefer one or the other. This special configuration was materialized with a thick vertical line in Figures 4 and 5. Irrespective of their pairwise correlation coefficients, the best combinations of substances appear as segments near the vertical, whereas alternatives are separated by sharp horizontal jumps. From these principles, Figures 4 and 5 showed, on average, four and five substances, respectively, in mixtures S1 and S2. The first four substances were truly present and only the fifth one, Irganox 1010, was falsely suggested in mixture S2. This false positive did not appear, because of its proximity with erucamide—they are very distant, as shown in Figure 2—and because of its analogy with other substances in the mixture: Irganox 1076 and Irganox 1035. It was remarkable that the ranking of substances remained unaffected by the presence of signals outside multiplet regions. The effects of nonzero baselines in noncoding regions were stronger in mixture S1, but remain almost insignificant above the optimal correlation threshold of 0.75.

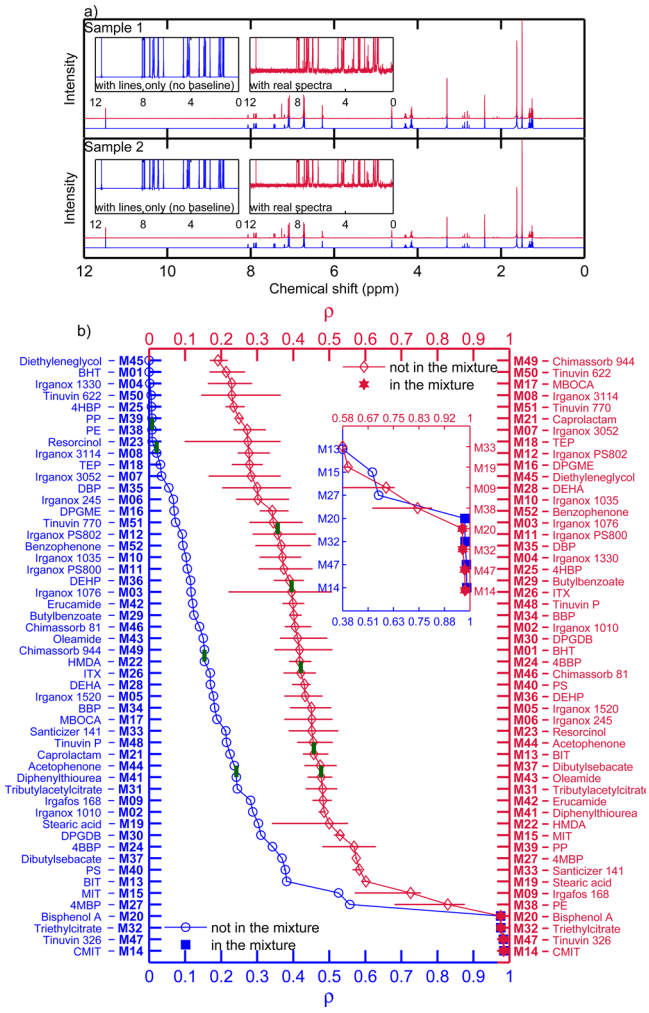


Figure 4. Substance ranking according to their similarity with mixture S1 based on a random mixing design involving 1000 samples (see Table 2): (a) two typical samples of mixture S1 combining peak lines (bottom curve) or real spectra (top curve) and (b) Pareto charts based on real spectra (right) and peak lines (left). 95% confidence intervals are plotted as horizontal lines. Thick vertical bars represent consecutive substances belonging to the same group of substances in Figure 2b at a threshold of 0.75. The insets show the details of each plot.

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

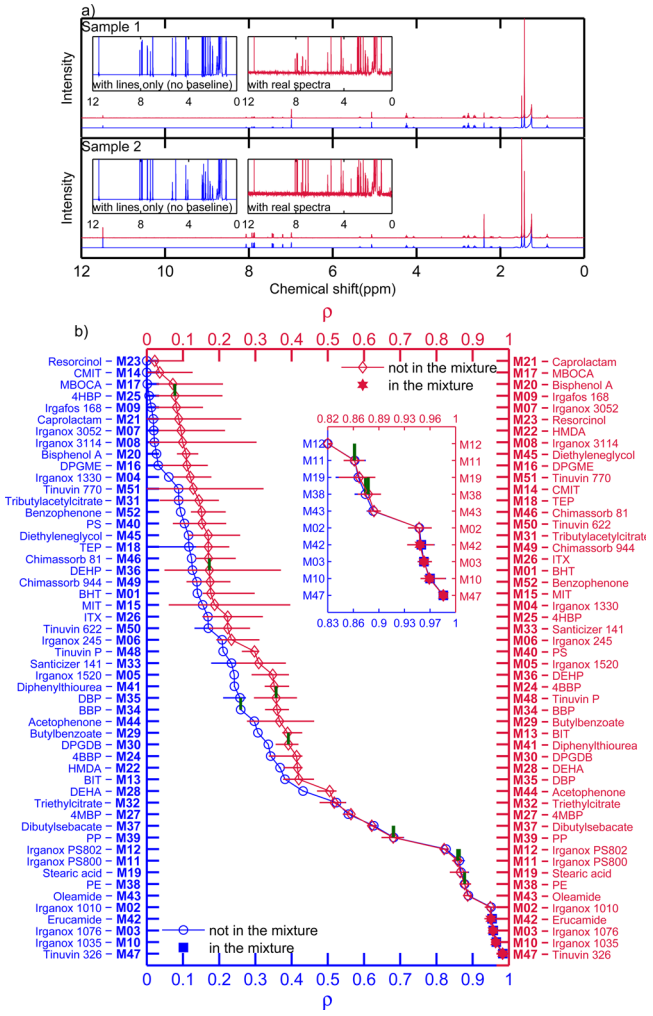


Figure 5. Substance ranking according to their similarity with mixture S2 based on a random mixing design involving 1000 samples (see Table 2): (a) two typical samples of mixture S2 combining peak lines (bottom curve) or real spectra (top curve) and (b) Pareto charts based on real spectra (right) and peak lines (left). 95% confidence intervals are plotted as horizontal lines. Thick vertical bars represent consecutive substances belonging to the same group of substances in Figure 2b at a threshold of 0.75. The insets show the details of each plot.

4.3. Deconvolution of Matching Cross-Correlation Functions. In mixtures S1 and S2 (Figures 4 and 5), the number of likely substances was mainly identified by the gap between pairwise correlation coefficients associated with best matching substances and intermediate matching ones. The gap was expected to occur above the threshold of 0.75, but without giving a clear decision rule when two or more gaps occurred. In addition, the scale invariance makes it only qualitative. A direct calculation of the proton molar fraction for each likely substance requires a full or partial deconvolution procedure. We proposed to carry out this operation in a least-squares sense (see eqs 7−9) by replacing previous correlation coefficients by cross-correlation functions. The principles are illustrated in Figures S3 and S4 in the Supporting Information[43] for two particular combinations of mixtures S1 and S2. Choosing seven substances instead of the full library reduced the size of the stiffness matrix $\mathbf{S}_{RR}$ from 279 × 52 down to 34 × 7 and 46 × 7, for mixtures S1 and S2, respectively. This operation can precomputed and only the choice of the considered test substances is dependent on the mixture. The load vector $\mathbf{S}_{mixture}$ is conversely built specifically for each mixture by collecting the 34 maxima of cross-correlation functions between multiplets and the spectrum each mixture. As shown in Figure S1 in the Supporting Information,[43] the signature of a given multiplet is identified visually, by using an odd number of lines distributed symmetrically around the position, giving the maximum of correlation, δ_{max}. This simple criterion allows us to discard Irgafos 168, Santicizer 141, and stearic acid from the possible list of substances of mixture S1 (their cross-correlation functions are asymmetric), whereas they exhibited pretty good pairwise correlations.

In the subsequent analysis, proportions of the four compounds in mixtures S1 and S2 (see Table 2), are random (10^3 samples), so that any of them was present in <10% w/w and in >90% w/w. The sparse deconvolution procedure based on eq 9 was initiated on the ten likeliest substances in Figures 4 and 5. The results are expressed as relative errors for substances truly present and absolute errors (proportion of false positives) for others. They are plotted against the number of inserted substances in Figures 6a and 6b, respectively. Baseline effects were added or not to study the specific effects of baseline shifts on false positives.

As expected from the pairwise correlation charts of Figures 4 and 5, mixture S1 was easier to resolve, mostly due to the quasi-absence of overlaps between substances. The relative error associated with truly present substances in mixture S1 did not exceed 14% for 95% of repetitions. The number of false positives was zero on average and appeared sparingly only for 2 substances ranked in fifth and sixth positions in Figure 4. Non-negativity constraints allowed the solution of mixture S1 to remain remarkably sparse and accurate with up to 10 substances. The performances were lower with mixture S2 but remained highly satisfactory. Oleamide and Irganox 1010—with spectra that were poorly separable from erucamide and Irganox 1076, respectively—were the two possible false positives and were observed in less than one-third of samples.

In previous numerical examples, the concentrations were random but chosen within the same range. Deconvolution errors were found within experimental errors. Nevertheless, such results cannot be extended to real samples as noises and biases caused by acquisition, numerical treatment, and, most of all, the presence of oligomers, degradation products, etc. in the extract. These specific effects are considered in mixtures S3−S7, which were prepared from real-life polymer extracts.

4.4. Deformulation of Real Materials. Since polyolefins and polyesters are prone to release significant amount of oligomers, their deformulation present significant complications by both conventional techniques (see Chapter 2 in ref 64) and rapid techniques. When the concentrations in additives are particularly low, such as in PLA, a prior precipitation step of oligomers is usually required.[65] We propose to apply ^{1}H NMR spectroscopy and our blind deconvolution procedure directly to the extracts of five materials (three polyolefins and two polyesters) without any physical separation. The collected spectra, denoted as S3−S7 in Table 2, were separated artificially into two categories, according to the mathematical difficulty. The deconvolution was said to be of "low" or "moderate" complexity, when one the following conditions was fulfilled:

- the number of substances in the mixture was low (e.g., two or less);

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

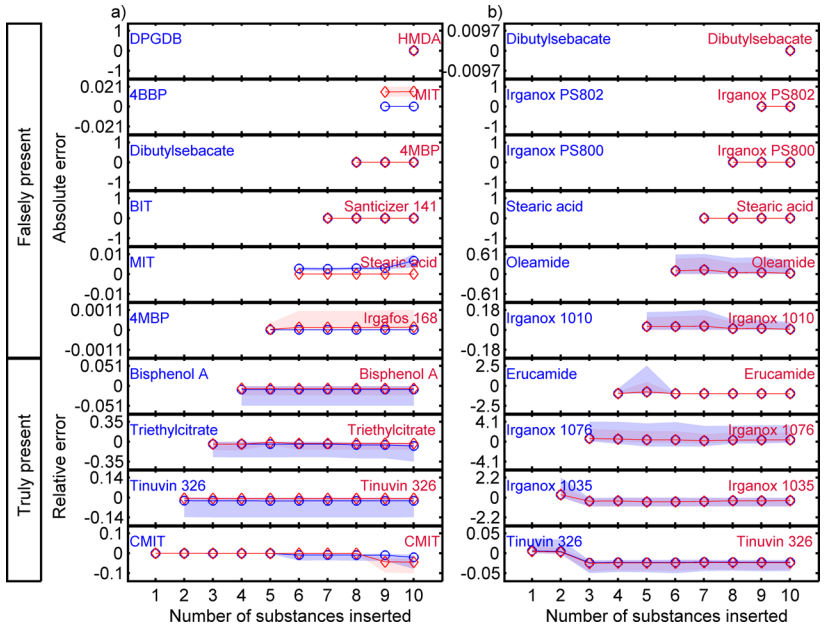
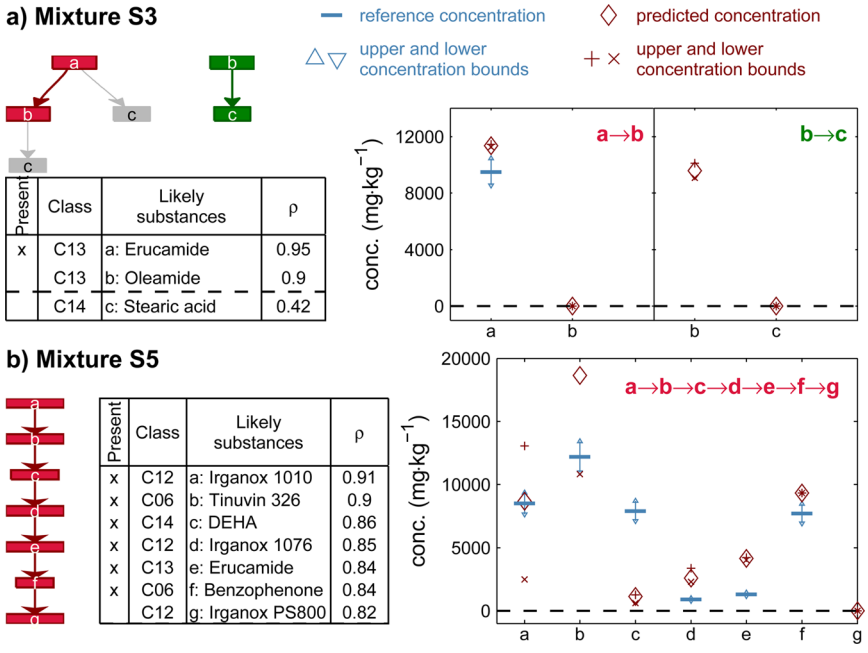


Figure 6. Deconvolution errors associated with mixtures (a) S1 and (b) S2 in random proportions (10^3 random samples), as a function of the number of substances in the non-negative least-squares system (9). Results associated with spectra with and without baselines are plotted with diamonds and circles. 95% confidence intervals are plotted as shaded areas.



Present	Class	Likely substances	ρ
x	C13	a: Erucamide	0.95
	C13	b: Oleamide	0.9
	C14	c: Stearic acid	0.42

Present	Class	Likely substances	ρ
x	C12	a: Irganox 1010	0.91
x	C06	b: Tinuvin 326	0.9
x	C14	c: DEHA	0.86
x	C12	d: Irganox 1076	0.85
x	C13	e: Erucamide	0.84
x	C06	f: Benzophenone	0.84
	C12	g: Irganox PS800	0.82

Figure 7. Predicted concentrations for extracts (a) S3 (two scenarios) and (b) S5 (one scenario), according to the insertion scenarios set by Schulze's rule. The substances are indexed with roman letters in the tables besides. The insertion scenarios appear as tree diagrams (colored nodes match likely scenarios read from root node). Concentration ranges depict the span of estimations with the number of inserted substances. The likeliest concentration was determined as the median of all determinations, while the condition number of $\mathbf{S}_{\mathbf{R}_{\text{choice}}\mathbf{R}_{\text{choice}}}$ was almost invariant with the number of substances.

- the substances (regardless their numbers) belong to different classes above a threshold correlation coefficient of 0.75 (as shown in Figure 2b);
- all the additives are within the library; and
- the residues after deconvolution (i.e., due to oligomers, unknown substances, baseline shifts, etc.) are small.

Samples S3, S5, and S7 fell in the first category and the results of their deconvolutions of their spectra are plotted in Figures 7a, 7b, and 8a, respectively. As they were extracted from different materials (PE, PP, and PLA, respectively), the deformulation complexity was not specifically associated with any material. The insertion scenarios were short or poorly branched. The number of identifiable substances (a close estimate of R_M) was also easy to guess from the increase of the condition number of $\mathbf{S}_{\mathbf{R}_{\text{choice}}\mathbf{R}_{\text{choice}}}$ or from the fitting error $\|\mathbf{S}_{\mathbf{R}_{\text{choice}}\mathbf{R}_{\text{choice}}}\mathbf{C} - \mathbf{S}_{\text{mixture}}\|$. Samples S4 and S6 belonged

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

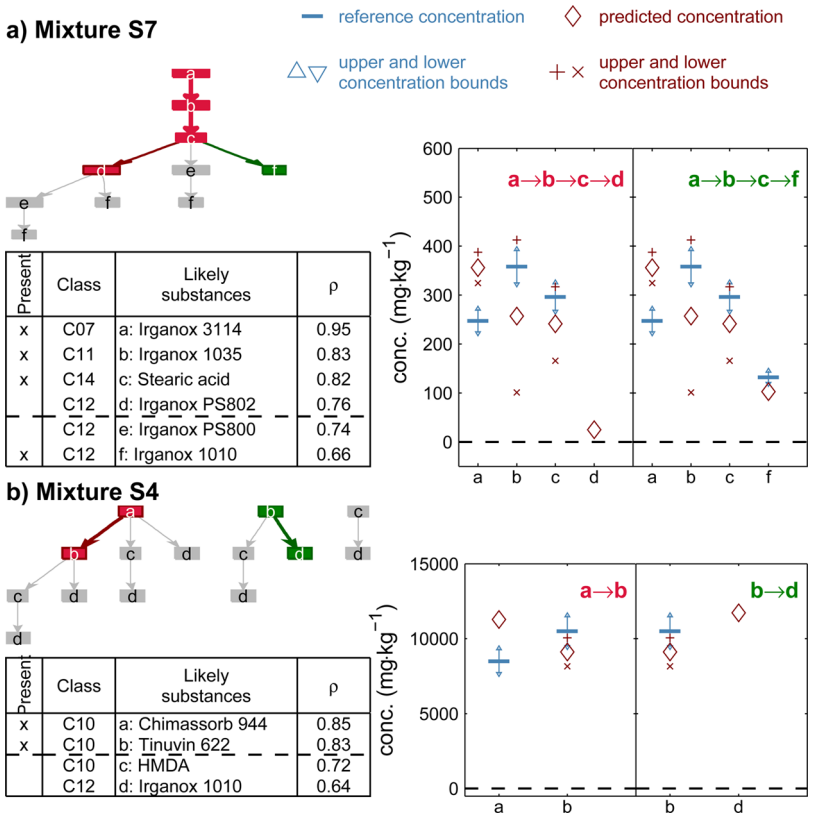

a) Mixture S7
reference concentration
predicted concentration
upper and lower concentration bounds
upper and lower concentration bounds
Present | Class | Likely substances | ρ
x | C07 | a: Irganox 3114 | 0.95
x | C11 | b: Irganox 1035 | 0.83
x | C14 | c: Stearic acid | 0.82
 | C12 | d: Irganox PS802 | 0.76
 | C12 | e: Irganox PS800 | 0.74
x | C12 | f: Irganox 1010 | 0.66
a→b→c→d
a→b→c→f
conc. (mg·kg⁻¹)
b) Mixture S4
Present | Class | Likely substances | ρ
x | C10 | a: Chimassorb 944 | 0.85
x | C10 | b: Tinuvin 622 | 0.83
 | C10 | c: HMDA | 0.72
 | C12 | d: Irganox 1010 | 0.64
a→b
b→d
conc. (mg·kg⁻¹)

Figure 8. Predicted concentrations for extracts (a) S7 (two scenarios) and (b) S4 (two scenarios), according to the insertion scenarios set by Schulze's rule.

conversely to the second category and led to multiple scenarios (considering 7 substances generated 534 and 518 nested scenarios, respectively) with multiple roots. Fortunately, transforming the original scenarios into preference lists collapsed the original tree into two main branches as shown in Figures 8b and 9, where non-negative solutions could be easily tracked.

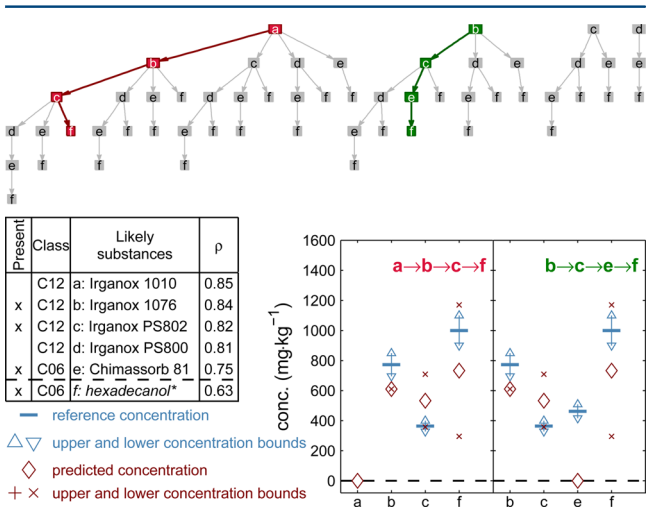

Present | Class | Likely substances | ρ
 | C12 | a: Irganox 1010 | 0.85
x | C12 | b: Irganox 1076 | 0.84
x | C12 | c: Irganox PS802 | 0.82
 | C12 | d: Irganox PS800 | 0.81
x | C06 | e: Chimassorb 81 | 0.75
x | C06 | f: hexadecanol* | 0.63
a→b→c→f
b→c→e→f
conc. (mg·kg⁻¹)
reference concentration
upper and lower concentration bounds
predicted concentration
upper and lower concentration bounds

Figure 9. Predicted concentrations for extract S6 (two scenarios) according to the insertion scenarios set by Schulze's rule. Asterisk (*) notes that data for this substance was based on the ^{1}H NMR spectrum of 1-hexadecanol, as calculated with ChemNMR from ChemBioDraw Ultra 14 Suite.[57]

Extract of sample S3 offers the simplest mixture with one single substance added to the polymer. It enabled us to test, in a blind manner, the ability to recognize the right substance in a real extract. The method correctly identified the substance **a** present with a very good pairwise correlation coefficient (ρ = 0.95). Because of the strong similarities between substances **a** (erucamide) and **b** (oleamide) (they belong to the same class C13 in Figure 2b), two major scenarios were constructed, starting from **a** (the likeliest) or from **b** (as an alternative to **a**). Spanning alternative scenarios enabled us to consider *a priori* the combination **a**+**b**, as well as **b** alone. When **a**+**b** was considered, **b** was not found after deconvolution (i.e., *a posteriori*) but **b** alone could replace **a** in all alternative deconvolutions. The equivalent amount of **b** was determined to be lower than that of **a**, because of a slightly lower correlation coefficient (ρ = 0.9). In addition, since substance **c** (stearic acid) belonged to a class (C14) that was closely related to that of **a** and **b** (C13; see Figure 2b), **c** was falsely detected each time it was inserted into the second position. Other substances were not significantly detected when scenarios were enlarged.

Similar to extract S3, extract S5 included six substances. They were all identified in a row (no alternative scenario) in Figure 7b, without significant risk of false positives. Since substance **a** (Irganox 1010) could be confused with a substance of the same class (**d**: Irganox 1076), the concentration range of **a** was broad. All substances were correctly quantified except substance **c** (bis(2-ethylhexyl)adipate), whose structure makes its spectrum very close to any aliphatic segment.

Extract S7 (see Figure 8a) led to more standard results with substances appearing in a decreasing order, relative to their importance in the extract. Alternative combinations appeared starting from the fourth inserted substance which could be

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

replaced by three substances belonging to the same class (C12): **d**, **e**, and **f**. Two of them, **d** and **e**, were false positives. The Schulze method eliminates substance *d* based on the concept of strongest path.[66] The two remaining substances were sorted as **d** > **f** and only the non-negative solutions of scenarios **a**+**b**+**c**+**d** and **a**+**b**+**c**+**f** enabled us to discriminate them. Concentrations of substance **d** was found very low, without being zero. All concentrations were correctly estimated except for substance **b** (Irganox 1035), whose groups present strong similarities with other Irganox substances (including either oxopropyl or *tert*-butyl groups) present in the list.

Extracts S4 (see Figure 8b) and S6 (see Figure 9) led many crossovers between classes, identified by multiple branching in insertion scenarios. For scenario S6, the sharp decay of singular values of $\mathbf{S}_{\mathbf{R}_{choice}\mathbf{R}_{choice}}$ beyond two substances confirmed that no more than substances could be reasonably identified. The Schulze preference list reduced the initial six scenarios of length 2 to two alternatives: **a**+**b** and **b**+**d**. By contrast with extract S3 (Figure 7a), the deconvolution did not help to choose between the two options. Substance **d** (Irganox 1010) was confused with **b** (Tinuvin 622). Substance **b** is indeed oligomeric amine light stabilizer with a very high molecular weight (3100−4000 g mol^{-1}) and numerous aliphatic protons.

Extract S6 introduced a substantial complication: 1-hexadecanol was a substance gathering ca. 28% of all protons (see Table 2), but which was not present in our library. In Figure 9, we present the results when the calculated spectrum of 1-hexadecanol is introduced, along the 52 experimental spectra of pure substances. From eq 12, it was guessed that no more than four substances could be identified. Two scenarios were identified and none of them gave the full answer. In the scenario **a**+**b**+**c**+**d**, **a** was finally correctly identified as a false positive but **3** (Chimasorb 81) was missing. In the scenario **b**+**c**+**e**+**f**, **e** was not identified either, because of a lack of nonaliphatic protons. From a practical point of view, the performances of the method were highly acceptable on this very complex mixture and when a nonexperimental spectrum was introduced. It is worth noticing that the removal of 1-hexadecanol from the list of possible substances led to one false positive: either Irganox PS800 or Irganox 1010.

5. CONCLUSIONS

Although the use of ^{1}H NMR spectrometers is not as widespread as the use of FTIR spectrometers on an industrial scale, the high specificity of NMR signal offers a better alternative to identify possible migrants from polymers in complex mixtures. When combined with appropriated deconvolution techniques, the proposed one-dimensional (1D) NMR spectroscopy may approach two-dimensional (2D) NMR spectroscopy performances for rapid identification and approximate quantification of polymer extracts. In this study, identification relies on an expandable library of spectra of pure substances. Without any loss of resolution, each spectrum was digitized without a baseline and spanned onto a finite basis of Gaussian kernels. The compression ratio reached 1:1000 or more while retaining the quantitative properties of the original signal. In particular, the proportionality of the cumulated signal with the number of protons was well-verified for each multiplet. To bring robustness, each multiplet signal was thought to be displaceable to follow possible peak shifts or peak broadenings in the spectrum of the unknown mixture. The proposed identification and quantification procedure generalizes conventional pursuit methods, designed originally to separate orthogonal signals, by introducing two scales of correlations. The global scale (entire spectrum) is used to set a consistent (but too large) list of substances of the library without excluding the possibility that unknown substances may be also present at arbitrary concentrations. The microscopic scale (multiplet scale) is considered to assign local correlations between details in the spectrum of the mixture and atoms inside the library. This last step feeds a least-squares formulation of the blind deconvolution problem. By combining both macroscopic and microscopic scales, a series of likely mixture scenarios with different numbers and combinations of substances is built. The entire system remains tractable by transforming all insertion scenarios into a preference list and by picking the substances as in a voting system. The entire approach was tested on both numeric examples and real material extracts. It provided remarkable capabilities of identification, regardless of the presence of oligomers and signal shifts. The low number of false positives makes it complementary of reference techniques, such as chromatographic and mass spectroscopy and hyphenated techniques (HPLC-MS, GC-MS, HPLC-NMR, FTIR-MS). The main envisioned applications are screening purposes, rapid detection of substances on positive/negative lists and maximum amounts in materials. It is indeed directly applicable for tailored identification and deformulation even when residues are large and cannot be explained with substances listed in the considered library.

Future works apply the methodology to the blind deformulation of 98 real food packaging materials on the French market along the project SAFEFOODPACK DESIGN.[67] Most of the mathematical developments and the library of coefficients are available as an open-source project so-called "SFPDnmrspec".[68]

■ ASSOCIATED CONTENT

Supporting Information

It contains one table (Table S1), four figures (Figures S1−S4), and four appendices (S1−S4). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: (+33) 169-935-063. Fax: (+33) 169-935-024. E-mail: olivier.vitrac@agroparistech.fr.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like thank Laurent Bélard (Biopolynov) and the Jeune Chambre Economique de Plasturgie for the processing of the reference materials used in this study. We thank the Institut de Chimie Moléculaire et des Matériaux d'Orsay (UMR 8182) and Pr. Denis Merlet for assistance in acquiring the reference NMR spectra. Furthermore, we would like to acknowledge the Agence Nationale de la Recherche (Project SAFEFOODPACK DESIGN, No. NR-10-ALIA-009) and the Association Nationale de la Recherche et de la Technologie (ANRT) for their financial support.

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

REFERENCES

(1) Corrigendum to Regulation (EC) No. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No. 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. *Off. J. Eur. Union* **2007**, *50*, 3.

(2) Regulation (EC) No 1935/2004 of the European parliament and of the council of 27 October 2004 on materials and articles intended to come into contact with food and repealing Directives 80/590/EEC and 89/109/EEC. *Off. J. Eur. Union* **2004**, *L338*, 4.

(3) Commission regulation (EU) No 10/2011 of 14 January 2011 on plastic materials and articles intended to come into contact with food. *Off. J. Eur. Union* **2011**, *L12*, 1.

(4) Commission regulation (EC) No 2023/2006 of 22 December 2006 on good manufacturing practice for materials and articles intended to come into contact with food. *Off. J. Eur. Union* **2006**, *L384*, 75.

(5) EuPIA. *Swiss Ordinance on Materials and Articles in Contact with Food (817.023.21)*; Federal Department of Home Affairs (FDHA): Bern, Switzerland, 2005.

(6) Commission directive 2004/1/EC of 6 January 2004 amending Directive 2002/72/EC as regards the suspension of the use of azodicarbonamide as blowing agent. *Off. J. Eur. Union* **2004**, *L7*, 45.

(7) Diamanti-Kandarakis, E.; Bourguignon, J.-P.; Giudice, L. C.; Hauser, R.; Prins, G. S.; Soto, A. M.; Zoeller, R. T.; Gore, A. C. Endocrine-Disrupting Chemicals: An Endocrine Society Scientific Statement. *Endocr. Rev.* **2009**, *30*, 293.

(8) Yang, C. Z.; Yaniger, S. I.; Jordan, V. C.; Klein, D. J.; Bittner, G. D. Most Plastic Products Release Estrogenic Chemicals: A Potential Health Problem That Can Be Solved. *Environ. Health Perspect.* **2011**, *119*, 989.

(9) Vandenberg, L. N.; Colborn, T.; Hayes, T. B.; Heindel, J. J.; Jacobs, D. R., Jr.; Lee, D.-H.; Soto, A. M.; vom Saal, F. S.; Welshons, W. V.; Zoeller, R. T.; Myers, J. P. Hormones and Endocrine-Disrupting Chemicals: Low-Dose Effects and Nonmonotonic Dose Responses. *Endocr. Rev.* **2012**, *33*, 378.

(10) Commission implementing regulation (EU) No 321/2011 of 1 April 2011 amending Regulation (EU) No 10/2011 as regards the restriction of use of Bisphenol A in plastic infant feeding bottles. *Off. J. Eur. Union* **2011**, *L87*, 1.

(11) FDA. The Food and Drug Administration is amending the food additive regulations to no longer provide for the use of polycarbonate (PC) resins in infant feeding bottles (baby bottles) and spill-proof cups, including their closures and lids, designed to help train babies and toddlers to drink from cups (sippy cups) because these uses have been abandoned. *Federal Registrer* **2012**, *77*, 41899.

(12) FDA. *FDA Regulations No Longer Authorize the Use of BPA in Infant Formula Packaging Based on Abandonment; Decision Not Based on Safety*. http://www.fda.gov/food/newsevents/constituentupdates/ucm360147.htm (accessed Dec. 29, 2014).

(13) FDA. *21 CFR 175: Indirect Food Additives: Adhesives and Components of Coatings*, 2013; pp 41840−41843.

(14) Loi No. 2012−1442 du 24 décembre 2012 visant à la suspension de la fabrication, de l'importation, de l'exportation et de la mise sur le marché de tout conditionnement à vocation alimentaire contenant du bisphénol A. *J. Off. Repub. Fr.* **2012**, *03000*, 20395.

(15) Geueke, B.; Wagner, C. C.; Muncke, J. Food Contact Substances and Chemicals of Concern: a Comparison of Inventories. *Food Addit. Contam., Part A* **2014**, *31*, 1438.

(16) Feigenbaum, A.; Scholler, D.; Bouquant, J.; Brigot, G.; Ferrier, D.; Franz, R.; Lillemark, L.; Riquet, A. M.; Petersen, J. H.; van Lierop, B.; Yagoubi, N. Safety and Quality of Food Contact Materials. Part 1: Evaluation of Analytical Strategies to Introduce Migration Testing into Good Manufacturing Practice. *Food Addit. Contam.* **2002**, *19*, 184.

(17) Castle, L. Assessment of consumer exposure to substances migrating from food contact materials. In *Final Conference of the EU project "FACET" (Flavours, Additives, and food Contact Materials Exposure Tool)*, Bach, A.; Simoneau, C., Eds. Brussels, Belgium, 2012; pp 93−100.

(18) *European Union Reference Laboratory for Food Contact Materials: Migration methods*. http://ihcp.jrc.ec.europa.eu/our_labs/eurl_food_c_m/migration-methods (accessed March 13, 2014).

(19) Potyrailo, R.; Rajan, K.; Stoewe, K.; Takeuchi, I.; Chisholm, B.; Lam, H. Combinatorial and High-Throughput Screening of Materials Libraries: Review of State of the Art. *ACS Comb. Sci.* **2011**, *13*, 579.

(20) *SpecInfo on the Internet—NMR, IR, MS*, http://onlinelibrary.wiley.com/book/10.1002/9780471692294 (accessed Dec. 29, 2014).

(21) ACD/Labs. Advanced Chemistry Development, Inc., http://www.acdlabs.com (accessed Dec. 29, 2014).

(22) CambridgeSoft Corp. http://www.cambridgesoft.com (accessed Dec. 29, 2014).

(23) Mestrelab Research. http://www.mestrelab.com (accessed Dec. 29, 2014).

(24) Modgraph Consultants. http://www.modgraph.co.uk (accessed Dec. 29, 2014).

(25) PERCH Solutions, Ltd. http://www.perchsolutions.com (accessed Dec. 29, 2014).

(26) Laatikainen, R.; Hassinen, T.; Lehtivarjo, J.; Tiainen, M.; Jungman, J.; Tynkkynen, T.; Korhonen, S. P.; Niemitz, M.; Poutiainen, P.; Jaaskelainen, O.; Vaisanen, T.; Weisell, J.; Soininen, P.; Laatikainen, P.; Martonen, H.; Tuppurainen, K. Comprehensive Strategy for Proton Chemical Shift Prediction: Linear Prediction with Nonlinear Corrections. *J. Chem. Inf. Model.* **2014**, *54*, 419.

(27) Bingol, K.; Brüschweiler, R. Deconvolution of Chemical Mixtures with High Complexity by NMR Consensus Trace Clustering. *Anal. Chem.* **2011**, *83*, 7412.

(28) Kwan, E. E.; Huang, S. G. Structural Elucidation with NMR Spectroscopy: Practical Strategies for Organic Chemists. *Eur. J. Org. Chem.* **2008**, *2008*, 2671.

(29) Pauli, G. F.; Jaki, B. U.; Lankin, D. C. Quantitative ^{1}H NMR: Development and Potential of a Method for Natural Products Analysis. *J. Nat. Prod.* **2004**, *68*, 133.

(30) Elyashberg, M. E.; Williams, A. J.; Martin, G. E. Computer-Assisted Structure Verification and Elucidation Tools in NMR-Based Structure Elucidation. *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, *53*, 1.

(31) Gillet, G.; Vitrac, O.; Desobry, S. A Fast Method to Assess the Composition of a Polyolefin: An Application to Compliance Testing of Food Contact Materials. *J. Appl. Polym. Sci.* **2011**, *119*, 1492.

(32) Schulze, M. A new monotonic, clone-independent, reversal symmetric, and condorcet-consistent single-winner election method. *Soc. Choice Welfare* **2011**, *36*, 267.

(33) Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37.

(34) Hyvarinen, A.; Oja, E. Independent Component Analysis: Algorithms and Applications. *Neural Networks* **2000**, *13*, 411.

(35) Toumi, I.; Caldarelli, S.; Torresani, B. A review of blind source separation in NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, *81*, 37.

(36) Naanaa, W.; Nuzillard, J.-M. Blind source separation of positive and partially correlated data. *Signal Process.* **2005**, *85*, 1711.

(37) Kopriva, I.; Jerić, I. Blind Separation of Analytes in Nuclear Magnetic Resonance Spectroscopy and Mass Spectrometry: Sparseness-Based Robust Multicomponent Analysis. *Anal. Chem.* **2010**, *82*, 1911.

(38) Kopriva, I.; Jeric, I.; Smrecki, V. Extraction of multiple pure component ^{1}H and ^{13}C NMR spectra from two mixtures: Novel solution obtained by sparse component analysis-based blind decomposition. *Anal. Chim. Acta* **2009**, *653*, 143.

(39) Sun, Y.; Xin, J. A Sparse Semi-Blind Source Identification Method and Its Application to Raman Spectroscopy for Explosives Detection. *Signal Process.* **2014**, *96* (Part B), 332.

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681

(40) Alonso, A.; Rodriguez, M. A.; Vinaixa, M.; Tortosa, R.; Correig, X.; Julia, A.; Marsal, S. Focus: a robust workflow for one-dimensional NMR spectral analysis. *Anal. Chem.* **2014**, *86*, 1160.
(41) Mallat, S. G.; Zhang, Z. F. Matching Pursuits with Time-Frequency Dictionaries. *IEEE Trans. Signal Process.* **1993**, *41*, 3397.
(42) Daubechies, I. Time Frequency Localization Operators—A Geometric Phase-Space Approach. *IEEE Trans. Inf. Theory* **1988**, *34*, 605.
(43) Nguyen, P.-M.; Lyathaud, C.; Vitrac, O. A Two-scale Pursuit Method for the Tailored Identification and Quantification of Unknown Polymer Additives and Contaminants by ^{1}H NMR.*Ind. Chem. Eng. Res.*201510.1021/ie503592z. (See the Supporting Information.)
(44) Sun, Y.; Xin, J. A Recursive Sparse Blind Source Separation Method and Its Application to Correlated Data in NMR Spectroscopy of Biofluids. *J. Sci. Comput.* **2012**, *51*, 733.
(45) Morris, G. A.; Barjat, H.; Home, T. J. Reference Deconvolution Methods. *Prog. Nucl. Magn. Reson. Spectrosc.* **1997**, *31*, 197.
(46) Barjat, H.; Morris, G. A.; Swanson, A. G.; Smart, S.; Williams, S. C. R. Reference Deconvolution Using Multiplet Reference Signals. *J. Magn. Reson., Ser. A* **1995**, *116*, 206.
(47) Lawson, C. L.; Hanson, R. J. *Solving Least Squares Problems*; Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, 1987; p 350.
(48) Chen, D.; Plemmons, R. J. Nonnegativity constraints in numerical analysis. In *The Birth of Numerical Analysis*; Bultheel, A., Cools, R., Eds.; World Scientific: Danvers, MA, 2009; pp 109−139.
(49) Bro, R.; De Jong, S. A Fast Non-Negativity-Constrained Least Squares Algorithm. *J. Chemom.* **1997**, *11*, 393.
(50) Van Benthem, M. H.; Keenan, M. R. Fast Algorithm for the Solution of Large-Scale Non-Negativity-Constrained Least Squares Problems. *J. Chemom.* **2004**, *18*, 441.
(51) Kim, D.; Sra, S.; Dhillon, I. S. *A New Projected Quasi-Newton Approach for the Non-negative Least Squares Problem*, Technical Report No. TR-06-54; The University of Texas: Austin, TX, 2007; p 16.
(52) Franc, V.; Hlavac, V.; Navara, M. Sequential Coordinate-Wise Algorithm for the Non-negative Least Squares Problem. In *CAIP 2005: Computer Analysis of Images and Patterns*, Gagalowicz, A., Philips, W., Eds.; Springer−Verlag: Berlin, 2005; pp 407−414.
(53) Bellavia, S.; Macconi, M.; Morini, B. An Interior Point Newton-like Method for Non-Negative Least-Squares Problems with Degenerate Solution. *Numer. Linear Algerbr.* **2006**, *13*, 825.
(54) Cantarella, J.; Piatek, M. Tsnnls: A solver for large sparse least squares problems with non-negative variables. *CoRR* 2004, vol. cs.MS/0408029.
(55) Parkes, D. C.; Xia, L. A complexity-of-strategic-behavior comparison between Schulze's rule and ranked pairs. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, AAAI-12*; AAAI Publications: Toronto, Ontario, Canada, 2012; pp 1429−1435.
(56) Tideman, T. N. Independence of clones as a criterion for voting rules. *Soc. Choice Welfare* **1987**, *4*, 185.
(57) ChemNMR Pro of ChemBioDraw Ultra 14 Suite (Perkin−Elmer). http://www.cambridgesoft.com/Ensemble_for_Chemistry/ChemBioDraw/ChemBioDrawUltra14Suite/ (accessed Feb. 22, 2015).
(58) Murphy, J. *Additives for Plastics Handbook*, 2nd Edition; Elsevier Advanced Technology: Oxford, U.K., 2001; p 484.
(59) Pritchard, G. *Plastics Additives*; Smithers Rapra Technology: Shrewsbury, U.K., 2005; p 208.
(60) Crompton, T. R. *Additive Migration from Plastics Into Foods: A Guide for the Analytical Chemist*; Smithers Rapra Technology Limited: Shrewsbury, U.K., 2007; p 324.
(61) Forrest, M. J. *Coatings and Inks for Food Contact Materials*. Smithers Rapra Publishing: Shrewsbury, 2007; p 132.
(62) Zweifel, H.; Maier, R. D.; Schiller, M. *Plastics Additives Handbook*; 6th ed.; Hanser Publications: Munich, 2009; p 1222.
(63) Gratia, A.; Merlet, D.; Ducruet, V.; Lyathaud, C. A comprehensive NMR methodology to assess the composition of biobased and biodegradable polymers in contact with foods. *Anal. Chim. Acta* **2015**, *853*, 477.
(64) Hummel, D. O. *Atlas of Plastics Additives: Analysis by Spectrometric Methods*; Springer−Verlag: Berlin, Heidelberg, 2002; p 537.
(65) Salazar, R.; Domenek, S.; Courgneau, C.; Ducruet, V. Plasticization of Poly(lactide) by Sorption of Volatile Organic Compounds at Low Concentration. *Polym. Degrad. Stab.* **2012**, *97*, 1871.
(66) Schulze, M. A New Monotonic and Clone-Independent Single-Winner Election Method. *Voting Matters* **2003**, *17*, 9.
(67) *INRA SafeFoodPack Design*. http://modmol.agroparistech.fr/SFPD/ (accessed July 5, 2011).
(68) Nguyen, P.-M.; Vitrac, O.; Lyathaud, C. ^{1}H NMR toolbox "SFPDnmrspec" to deformulate plastic materials. https://github.com/ovitrac/SFPDnmrspec (accessed Dec. 30, 2014).

DOI: 10.1021/ie503592z
Ind. Eng. Chem. Res. 2015, 54, 2667−2681